

Construction Morphology and the interaction of syntax and word formation*

Geert Booij (University of Leiden)

1. Introduction

The relation between word formation and syntax has been a persistent topic of research in the publications of my esteemed colleague professor Soledad Varela. Therefore, I hope it is a proper tribute to her work on morphology to deal with certain aspects of this kind of interaction in my contribution to this volume on the occasion of her retirement.

Lexical units can be constructed in two ways, by means of word formation (morphological operations) and by syntax, as also pointed out in Rainer & Varela (1992) in their discussion of the demarcation of compounds and phrases in Spanish. One obvious function of word formation mechanisms such as compounding and derivation is extending the set of lexical units of a language. However, the set of lexical units of a language is larger than the set of words. Phrasal units of various size form lexical units as well, varying from small phrases to complete sentences, and the construction of such units contribute to the growth and the size of the lexicon as well. It is therefore important not to identify the notions 'word' and 'lexical unit'. Simplex words are lexical units by definition, and complex words are lexical units when they are conventionalized ('existing complex words'). Many complex words never make it to the status of lexical unit since they are not conventionalized, and are used only once. Hence, expanding the lexicon is

only one function of word formation. Its other function is to provide relatively compact linguistic expressions for denoting all kinds of concepts.

In this paper, I will discuss various types of interaction between word formation and syntax. In particular, I will show that paradigmatic relationships between phrasal and morphological schemas for the construction of lexical units play an important role. In section 2, I summarize the evidence for phrasal lexical units, a notion that we need for a proper analysis of the interaction of syntax and word formation in the construction of lexical units. Section 3 presents some basic notions of Construction Grammar that are presupposed in my analysis. The following two sections discuss a number of specific cases of morphology-syntax interaction. I will argue that these phenomena provide evidence for an architecture of the grammar in which there is no strict division between grammar and lexicon, and in which networks of paradigmatic relations between linguistic expressions play a crucial role in a proper account of the form and meaning of lexical units.

2. The pervasiveness of prefabs

In the linguistic literature lexical phrasal units are referred to as phraseologisms (Fleischer 1992; 1997) or prefabs (Erman & Warren 2000). A well known class of such lexical phrases are combinations of an adjective and a noun. Compare the following A+N phrases in English and their glosses in Dutch:

(1)	<i>English</i>	<i>Dutch</i>
-----	----------------	--------------

a.	strong tea / *mighty tea	sterke thee ‘strong tea’
	weak tea / *feeble tea	slappe thee ‘lit. slack tea’
b.	a steady girlfriend	een vaste vriendin ‘lit. a fixed girlfriend’
	a confirmed bachelor	een verstokte vrijgezel ‘lit. a hardened bachelor’
	an eligible bachelor	een begerenswaardige vrijgezel ‘lit. a desirable bachelor’

The English examples in (1a) are given by Erman and Warren (2000), who point out that there is a fixed choice of adjective in the examples (1a): neither can *strong* be replaced with *mighty*, nor *weak* with *feeble*. The Dutch glosses exhibit the same type of conventionality. In the case of *strong tea*, Dutch requires the direct translation of *strong*, the adjective *sterk*, but in the case of *weak tea*, the direct translation *zwak* cannot be used, and instead the adjective *slap* ‘loose’ is the proper choice. The same observation can be made for the English phrases and their Dutch glosses in (1b). Similarly, in a detailed study of English verbs, Dąbrowska concluded that “speakers have very specific knowledge about the collocations and semantic preferences of individual verbs” (Dąbrowska 2009).

The examples (1) illustrate that knowledge of the lexical units of a language comprises much more than knowledge of its vocabulary. Knowledge of conventional word combinations is essential for a full mastery of a language. In using a non-native language, we might speak grammatically correct, but un-idiomatically. Acquiring the competence of idiomaticity is therefore an important goal of second language education

(Boers et al. 2006). Special dictionaries of word combinations, such as Benson et al. (1997) for English, are very useful for achieving this goal.

The term ‘prefab’ used by Erman and Warren (2000) for this type of lexical unit is more adequate than the term ‘idiom’, because we tend to associate the latter term with the idea of non-compositionality: an idiom is usually assumed to possess a (partially) unpredictable meaning. The term ‘fixed expression’ is a good alternative (Sprenger 2003). The term ‘prefab’ rightly indicates that we have to do here with ready-made multi-word chunks that reflect the conventionality of language use. For example, there is no semantic opacity involved in the meaning of the English phrase *strong tea* (compare this phrase to the A+N phrases *red army* and *yellow fever* that do have idiomatic meanings). The specific knowledge involved in the case of *strong tea* is only that of the selection of the adjective *strong* rather than *mighty*. These kinds of collocations can therefore be qualified as ‘idioms of encoding’ (Gagné & Spalding 2006) since they do not pose a problem for decoding their meaning, but specify how a particular meaning should be encoded.

Knowledge of prefabs should be an important component of models of language production and perception (Kuiper et al. 2007; Sprenger 2003; Sprenger et al. 2006). Erman and Warren (2000) estimate that about 55 % of the utterances in written English consist of prefabs. Prefab knowledge is very important for language production because it is crucial in the task of lexical choice as part of the encoding process. Psycholinguistic evidence confirms the psychological reality and relevance of prefabs since phrasal sequences exhibit frequency effects, just like words, even when these phrases are compositional (Tremblay & Baayen 2010).

Knowledge of prefabs has received renewed attention in the linguistic literature, because its implications for our view of the architecture of the grammar, and of language change (Booij 2010; Bybee & Beckner 2010; Bybee & Torres Cacoullos 2009; Ernestus & Baayen 2011; Jackendoff 1997; 2002a; 2002b; Wray 2002; 2008).

Below, I will argue that the theory of Construction Grammar (CxG) and its application to the construction of lexical units, Construction Morphology (CM), provide an insightful account of the role and position of lexical units, including prefabs, in the grammars of natural languages.

3. Construction Grammar

In CxG both complex words and phrasal lexical units are constructs, that is, pairings of forms and meanings. Sets of linguistic constructs with corresponding, recurrent properties can be characterized by means of abstract schemas. Consider the English A+N phrases *strong tea* and *red army* mentioned above. They have the regular structure [A N]_{NP}, and hence are instantiations of the English Noun Phrase. A central idea of CxG is that both schemas and their instantiations are stored in the grammar. An abstract schema dominates its individual instantiations, which inherit by the properties defined in the schema. Thus, the predictable properties of individual AN phrases count as redundant information. This makes it possible to store completely regular but conventionalized phrases in the lexicon without obliterating their regular character.³ This view is also defended by Jackendoff in his Parallel Architecture (PA) approach to grammar (Jackendoff 2002a; 2010; 2011).

There are two main reasons for storing specific phrasal patterns. One reason is that they are constructions in the sense of (Goldberg 1995: 4), that is, they may have semantic properties that cannot be predicted from the meaning contribution of the constituent words. The other reason is that instantiations of phrasal patterns may be entrenched; therefore, we need a phrasal schema that generalizes over the common properties of entrenched phrases such as AN phrases (Goldberg 2006: 5), Zeschel (2009). Jackendoff (2010: 224) argues that we should allow for schemas that do not have a specific constructional meaning, such as the schemas for English AN phrases, particle verbs, and NN compounds. One may, however, maintain the position that what is stored are always constructions, since regular phrasal patterns can always be assigned a very general constructional meaning. In the case of English AN phrases, for instance, the constructional meaning is that of ‘modifier – modified’. That is, the construction imposes a modifier interpretation on the adjective, and takes the head noun to denote the modified concept.

In the remainder of this article, I will therefore assume the correctness of the assumption that both abstract schemas for lexical (syntactic and morphological) units, and their conventionalized instantiations are stored in the grammar. This can be modeled as a hierarchical lexicon, with constructions at different levels of expressions (Booij 2010). For Dutch compounds, for instance, we need sub-schemas because constituents of compounds may feature a specific ‘bound’ meaning when embedded in compounds. An example of this phenomenon is the set of words with intensifying meaning that occur as the left constituents of Dutch XA compounds, where X = N, A or V:

(2) Intensifying lexemes in Dutch X A compounds

X = noun

examples

ber-e 'bear'

bere-sterk 'very strong', bere-gezellig 'very cosy'

bloed 'blood'

bloed-serieus 'very serious', bloed-link 'very risky'

X = adjective

dol 'mad'

dol-blij 'very happy', dol-gelukkig 'very happy'

stom 'stupid'

stom-toevallig 'completely coincidental', stom-verbaasd
'very surprised'

X = verb

kots 'vomit'

kots-misselijk 'very sick', kots-beu 'very tired of'

loei 'sizzle'

loei-hard 'very hard', loei-goed 'very good'

In these words, the first constituent no longer carries its lexical meaning as specified in the left column, but functions as a word with an abstract intensifier meaning. In a few cases there might still be a semantic explanation for the intensifier interpretation of the modifying noun and the head: *bere-sterk*, for instance, can be interpreted as 'as strong as a bear'. However, there is no obvious way in which the word *gezellig* 'cosy' can be related to a typical property of bears. This semantic reanalysis of such words as intensifiers is made overt by the fact that they can be attached productively with this specific intensifier meaning to form new adjectives. A noteworthy point concerning the intensifiers *bere-*, *rete-*, and *reuze-* is that they consist of a consonant-final stem followed by a linking element *-e* [ə]. They appear without schwa when used as independent words. This linking element is a necessary part of these nouns when used as intensifiers.

We can formally express the affixoid nature of these compound-initial lexemes by specifying them in constructional idioms of the following form:

- (3) $[[\text{bere}]_{\text{Nk}} [\text{x}]_{\text{Ai}}]_{\text{Aj}} \leftrightarrow [\text{very}_k \text{SEM}_i]_j$
 $[[\text{dol}]_{\text{Ak}} [\text{x}]_{\text{Ai}}]_{\text{Aj}} \leftrightarrow [\text{very}_k \text{SEM}_i]_j$
 $[[\text{loei}]_{\text{Vk}} [\text{x}]_{\text{Ai}}]_{\text{Aj}} \leftrightarrow [\text{very}_k \text{SEM}_i]_j$

The symbol \leftrightarrow stands for the correspondence relation between form and meaning. On the left hand side we have a form specification, and on the right hand side a specification of the corresponding meaning. The semantic contribution of specific formal sub-constituents is indicated by means of co-indexation (Jackendoff 2002a). Constructional idioms are morphological or syntactic schemas in which one or more positions are lexically fixed, whereas other positions are open slots, represented by variables (Jackendoff 2002a). Being embedded in constructional schemas makes these words similar to affixes. The difference is that affixes do not carry a lexical category label, and hence cannot be related to independent lexemes in the lexicon. The schemas (3) specify that the words on the left do not have their regular lexical meaning but evoke, in these morphological constructions, the meaning of intensification ('to a high degree') or excellence.

4. Lexical phrases and names

In probably all European languages we find phrases with coordination that are semantically transparent, but have a fixed order, such as English *salt and pepper*, *father*

and son, ladies and gentlemen (Malkiel 1959). These conventionalized expressions have to be listed because of the fixed order in which the two words appear. On the other hand, they instantiate regular and productive phrasal patterns, and we should express this in our description. Therefore, the syntactic coordination schema $[N \text{ and } N]_{NP}$, which is a sub-case of English coordination, dominates its instantiations, phrases such as *salt and pepper*. The only non-redundant information concerning these *N and N* phrases is that they exist (that is, are conventionalized), and the order in which the two words appear.

Thus, there is no principled difference between complex words and phrases with respect to the division of labour between storage and computation. Storage of complex words is necessary, one reason being that we have to specify which possible complex words are conventional, that is actual words.

A correct prediction of the position that both words and phrases are stored as lexical units is that we will find competition and blocking effects between synonymous phrases and words. For instance, the existence of the Dutch AN compound *zuur-kool* ‘sour-cabbage, Sauerkraut’ impedes using the AN phrase *zure kool* ‘sour cabbage’ for this type of cabbage, and inversely, the existence of the AN phrase *rode kool* ‘red cabbage’ impedes the coinage of the AN compound *rood-kool*. Related languages may differ in the choice between synonymous but structurally different options. For instance, where German has a very productive pattern of AN compounds, Dutch speakers tend to prefer AN phrases (as in German *Rotwein* ‘red wine’ versus Dutch *rode wijn* ‘red wine’) (Booij 2002). Romance languages also use a lot of NA phrases as names for concepts.

Phrasal lexical units may also arise through a special form of syntax. For instance, names for functions, organizations, etc. may exhibit ‘headline syntax’, the type of syntax

that is allowed in the headlines of newspapers, with omission of grammatical words, as illustrated by the following example:

(4) director cultural affairs

In this construct, the head *director* is on the left. Hence it cannot be a compound, since English compounds are right-headed. On the other hand, it is not a regular phrase since in NPs the complement must be governed by a preposition, which is lacking here. Hence, it is a case of special syntax, allowed in headlines of newspapers, but also in names for functions (Booij 2009b). This implies the assumption of special phrasal schemas for extending the English fund of lexical expressions for functions and institutions:

(5) [$\dots N_i NP_j$]_{NPK} \leftrightarrow [SEM_i with relation R to SEM_j]_k SEM_i = function, institution, ...

5. Form-meaning asymmetries

As we saw above, phrases may form building blocks of complex words. But what happens if a word formation process does not take phrases as bases for affixation?

Consider the following examples from Italian:

(6)	chitarra elettrica ‘electric guitar’	chitarrista elettrico ‘electrical guitarist’
	flauto barocco ‘baroque flute’	flautista barocco ‘baroque flute player’
	tennis da tavola ‘tennis table’	tennista da tavolo ‘table tennis player’

The word sequences on the left and the right are NPs of the form A + N or N Prep N. The phrase *chitarrista elettrico* denotes someone who plays the electric guitar. However, the phrase *chitarra elettrica* ‘lit. guitar electric, electric guitar’ is not the formal base for the attachment of the suffix *-ista* that is used in Italian to create personal names. This is clear from the word order (the suffix is not attached at the right edge, **chitarra elettricista*), and from the fact that the adjective *elettrico* agrees in number and gender with the masculine noun *chitarrista*, and not with the feminine noun *chitarra*. This asymmetry between form and meaning (‘bracketing paradox’) is the same as that in the famous English example *transformational grammarian* discussed in Spencer (1988). In the case of English, one might consider the suffix *-ian* to be formally attached to the phrase *transformational grammar*, an option that is available due to the English word order and the poor inflection of English. Such an analysis is obviously impossible in the case of Italian, which only allows for a phrasal interpretation:

(7) [[chitarr-ista]_N [elettrico]_A]_{NP}

This suggests that the adjective *elettrico* has semantic scope over a part of the complex phrasal head *chitarrista*, and that the internal morphological structure of the modified head word is accessible. In example (7), the adjective *elettrico* modifies the nominal base of the head noun. In the following examples AN phrases used as modifiers are turned into AA sequences, in order to fit the canonical phrase structure of Italian (examples provided to me by Daniele Vergilitto):

6. Conclusions

In this article it was shown that syntax and morphology may interact in a number of ways in order to expand the set of names for concepts in a language. The naming function is not the exclusive domain of word formation. Phrases also function as names, and in addition they form bases from which names can be derived in special ways such as the formation of acronyms and stump compounds.

The use of phrases as units of name formation is increased by languages allowing for asymmetries between formal structure and semantic structure. Such asymmetries appeared not to form a problem in encoding or decoding, thanks to the availability of paradigmatic relationships between linguistic expressions.

These observations about the various relationships between phrases and complex words that play a role in constructing and expanding the lexicon of a language find a natural and insightful account in the theory of Construction Grammar, which makes use of phrasal and morphological schemas. These schemas can be unified, thus giving rise to complex words with phrases as building blocks. By specifying both abstract schemas and their instantiations, we can do justice to both the creative and the conventional aspect of linguistic knowledge. Moreover, schemas can be related paradigmatically, which provides the flexibility in coining lexical units that is needed.

*The research for this paper has been supported by a fellowship of the Alexander von Humboldt Foundation, which is gratefully acknowledged here.

References

- Benson, Morton, Evelyn Benson & Robert Ilson. 1997. The BBI dictionary of English word combinations Amsterdam / Philadelphia: Benjamins.
- Boers, Frank, June Eyckmans, Jenny Kappel, H el ene Stengers & Murielle Demecheleer. 2006. Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research* 10.245-61.
- Booij, Geert. 2010. *Construction morphology* Oxford: Oxford University Press.
- Bybee, Joan & Clay Beckner. 2010. Usage-based theory. *The Oxford handbook of linguistic analysis*, ed. by B. Heine & H. Narrog, 827-56. Oxford: Oxford University Press.
- Bybee, Joan & Rena Torres Cacoullos. 2009. The role of prefabs in grammaticization. How the particular and the general interact in language change. *Formulaic language. Vol 1. Distribution and historical change*, ed. by R.L. Corrigan, E. Moravcsik, H. Ouali & K.M. Wheatly, 195-217. Amsterdam / Philadelphia: Benjamins.
- Dąbrowska, Ewa. 2009. Words as constructions. *New directions in Cognitive Linguistics*, ed. by V. Evans & S. Pourcel, 201-23. Amsterdam / Philadelphia: Benjamins.
- Erman, Britt & Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text* 20.29-62.
- Ernestus, Mirjam & R. Harald Baayen. 2011. Corpora and exemplars in phonology. *Handbook of phonology*, ed. by A. Yu & J. Goldsmith. Oxford: Blackwell.

- Fleischer, Wolfgang. 1992. Konvergenz und Divergenz von Wortbildung und Phraseologisierung. *Phraseologie und Wortbildung. Aspekte der Lexikonerweiterung*, ed. by J. Korhonen, 53-65. Tübingen: Niemeyer.
- Fleischer, Wolfgang. 1997. *Phraseologie der deutschen Gegenwartssprache*. 2e, erweiterte Auflage Tübingen: Niemeyer.
- Gagné, Christine L. & Thomas L. Spalding. 2006. Conceptual combination: implication for the mental lexicon. *The representation and processing of compound words*, ed. by G. Libben & G. Jarema, 145-68. Oxford: Oxford University Press.
- Goldberg, Adele. 1995. *Constructions. A Construction Grammar approach to argument structure* Chicago: Chicago University Press.
- Goldberg, Adele. 2006. *Constructions at work. The nature of generalization in language* Oxford: Oxford University Press.
- Jackendoff, Ray. 1997. *The architecture of the language faculty* Cambridge Mass.: MIT Press.
- Jackendoff, Ray. 2002a. *Foundations of language* Oxford: Oxford University Press.
- Jackendoff, Ray. 2002b. What's in the lexicon? Storage and computation in the language faculty, ed. by S. Nootboom, F. Weerman & F. Wijnen, 3-40. Dordrecht: Kluwer.
- Jackendoff, Ray. 2010. *Meaning and the lexicon. The parallel architecture 1975-2010* Oxford: Oxford University Press.
- Jackendoff, Ray. 2011. Formal CxG: Constructions in the Parallel Architecture. *The Oxford handbook of Construction Grammar*, ed. by T. Hoffmann & G. Trousdale. Oxford: Oxford University Press.

- Kuiper, Koenraad, Marie-Elaine Van Egmond, Gerard Kempen & Simone Sprenger. 2007. Slipping on superlemmas. Multi-word lexical items in speech production. *The Mental Lexicon* 2.313-57.
- Rainer, Franz & Soledad Varela. 1992. Compounding in Spanish. *Rivista di Linguistica* 4.117-42.
- Sprenger, Simone A. 2003. Fixed expressions and the production of idioms Nijmegen: Max-Planck-Institut für Psycholinguistik.
- Sprenger, Simone, Willem J. M. Levelt & Gerard Kempen. 2006. Lexical access during the production of idiomatic phrases. *Journal of Memory and Language* 54.161-84.
- Tremblay, A. & R. Harald Baayen. 2010. Holistic processing of regular four-word sequences: A behavioral and erp study of the effects of structure, frequency, and probability on immediate free recall. *Perspectives on formulaic language. Acquisition and communication*, ed. by D. Wood, 151-73. London: The Continuum International Publishing Group.
- Wray, Alison. 2002. *Formulaic language and the lexicon* Cambridge: Cambridge University Press.
- Wray, Alison. 2008. *Formulaic language: pushing the boundaries* Oxford: Oxford University Press.
- Zeschel, Arne. 2009. What's (in) a construction? Complete inheritance vs. full-entry models. *New directions in Cognitive Linguistics*, ed. by V. Evans & S. Pourcel, 185-200. Amsterdam / Philadelphia: Benjamins.